(12)

AD A116582

SOME ASPECTS OF INFERENCE

FOR MULTIVARIATE INFINITELY DIVISIBLE DISTRIBUTIONS

by

STANLEY L. SCLOVE

TECHNICAL REPORT NO. 82-3

June 15, 1982

Accepted for publication in the journal STATISTICS & DECISIONS

PREPARED FOR THE
OFFICE OF NAVAL RESEARCH
UNDER
CONTRACT N00014-80-C-0408,
TASK NR042-443
with the University of Illinois at Chicago Circle

Principal Investigator: Stanley L. Sclove



DEPARTMENT OF QUANTITATIVE METHODS
COLLEGE OF BUSINESS ADMINISTRATION
UNIVERSITY OF ILLINOIS AT CHICAGO CIRCLE
CHICAGO, ILLINOIS 60680

SOME ASPECTS OF INFERENCE FOR
MULTIVARIATE INFINITELY DIVISIBLE DISTRIBUTIONS

Stanley L. Sclove

Departments of Mathematics and Quantitative Methods
University of Illinois at Chicago Circle

CONTENTS





A

SOME ASPECTS OF INFERENCE FOR
MULTIVARIATE INFINITELY DIVISIBLE DISTRIBUTIONS

Stanley L. Sclove

ABSTRACT

Measurement of dependence in the infinitely divisible class of multivariate distributions, based on developments in probability theory for that class, is discussed. It has been shown that pairwise independence is equivalent to mutual independence in this class. When the infinitely divisible variables contain no normal component (in particular, when they are discrete), the cumulant of order (2,2) can be used as a measure of pairwise dependence; when a normal component is present, the appropriate measure also involves the covariance. Results for testing independence of infinitely divisible random variables are discussed. A method of testing normality against infinitely divisible alternatives is given.

## 1. INTRODUCTION

A random variable (r.v.) X is infinitely divisible (inf.div.) if there exists a triangular sequence $X_{nj}$, $n = 1, 2, \ldots$, $j = 1, 2, \ldots, n$, such that, for each $n = 1, 2, \ldots$, the n r.v.'s $X_{nj}$, $j = 1, 2, \ldots, n$, are independent and identically distributed and the variables $X^{(n)}$ defined by $X^{(n)} = X_{n1} + X_{n2} + \ldots + X_{nn}$, $n = 1, 2, \ldots$, all have the same distribution as X. The condition in terms of the characteristic function (c.f.) $\phi(u)$ of the r.v. X is that, for each n there exist a c.f. $\phi_n(u)$ such that $\phi_n(u) = [\phi_n(u)]^n$. The notion of infinite divisibility applies whenever a notion of addition is defined; e.g., the variable X may be a vector or a matrix.

---

AMS classification: 62H15, 62H20

Key words and phrases: infinite divisibility; multivariate normality; kurtosis; measures of dependence

The marginal distributions of an inf.div. multivariate distribution are inf.div. I.e., if a random vector (r.vec.) $\underline{X} = (X_1, X_2, \ldots, X_p)'$ is inf.div., then each variable $X_v$ is inf.div., $v = 1, 2, \ldots, p$. Also, if each variable of a r.vec. is inf.div. and the variables are independent, then the r.vec. is inf.div. But the elements of inf.div. r.vec.'s can be dependent; so the class of multivariate inf.div. distributions is quite broad. In particular, the class is closed under affine transformations.

Using the canonical representation of the c.f. of inf.div. r.vec.'s (see, e.g., [12] and the discussion below), a necessary and sufficient condition for mutual independence of the elements of the r.vec. was obtained in [9]. When the means are zero, this condition is simply that the squares of the variables be uncorrelated. Thus, for inf.div. r.vec.'s, not only does mutual independence reduce to pairwise independence but also the parametric characterization of dependence is simple. In the present paper some methods of statistical inference based on this advance in probability theory are developed.

It is not too much of an over-simplification to say that the applicability and relative ease of implementation of procedures derived from multivariate normal distributions depends upon the propriety of the correlation coefficient as the measure of dependence between variables which are jointly normally distributed. As will be discussed below, the development in [9] includes a measure of dependence for the variables of inf.div. r.vec.'s. The class of multivariate inf.div. distributions includes the multivariate normal family as well as other continuous multivariate distributions. Further, it includes discrete multivariate distributions, useful for modeling data such as that generated by multivariate point processes. Finally, and perhaps most importantly, the class includes r.vec.'s which are mixed in the sense that some variables are continuous and others are discrete. Here are some examples of sources of such data, to mention just two. One is in the observation of physical systems where one considers continuous measurements of energy,

phase, angular momentum, together with counts of numbers of collisions, disintegrations, etc. Another source is in the observation of biomedical systems where one considers continuous measurements of blood pressures, pulse rates, chemical concentrations, etc., together with counts of red and white blood cells. The measures of dependence stemming from [9] provide a method for the systematic treatment of dependence among the variables of such mixed r.vec.'s. These measures of dependence are potentially of great importance and applicability, as they play a role analogous to that of correlation coefficients.

Multivariate inference problems assuming inf.div. distributions are relatively tractable. Results demonstrating this are given below. Sections 2 and 3 present some results for inf.div. probability laws; the results and the proofs are at least implicit in [9]. These results form the basis for the remarks on inference of Section 4.

Now suppose $\underline{X}$ is an inf.div. r.vec. and let $\phi(\underline{u})$ be its c.f. The Kolmogorov representation for the logarithm $\psi(\underline{u})$ of $\phi(\underline{u})$ is

$$(1.1) \qquad \psi(\underline{u}) = i\underline{u}'\underline{\mu} - u'\underline{\Gamma}\,\underline{u}/2 + \int[\exp(i\underline{u}'x)-1-i\underline{u}'\underline{x}](\underline{x}'\underline{x})^{-1}M(d\underline{x}),$$

where M is a bounded measure having no mass at the origin and $\underline{\Gamma}$ is a positive definite matrix. Let $U \equiv V$ mean that U and V have the same distribution. The representation (1.1) means that

$$(1.2) \qquad \underline{X} \equiv \underline{X}_G + \underline{X}_P,$$

where $\underline{X}_G$ and $\underline{X}_P$ are independent, the component $\underline{X}_G$ is of Gaussian type, i.e., has a multivariate normal distribution, and the component $\underline{X}_P$ is of Poisson type, i.e., has log c.f. equal to the integral in (1.1).

## 2. TWO VARIABLES

Now consider two jointly inf.div. r.v.'s X and Y, i.e., let (X,Y)' be an inf.div. r.vec.

## 2.1. Characteristic function

The Kolmogorov representation (1.1) holds in the case of finite variances and for two variables is of the form

$$(2.1) \qquad \psi(t,u) = it\mu_1 + iu\mu_2 - t^2\gamma_{11}/2 - tu\gamma_{12} - u^2\gamma_{22}/2$$

$$+ \int\{\exp[i(tx + uy)] - 1 - i(tx + uy)\}\,(x^2 + y^2)^{-1}M(dx,dy).$$

## 2.2. Cumulants

The cumulants $\kappa_{rs}$, when they exist, are defined by the expansion

$$\psi(t,u) = \sum \sum \kappa_{rs}\, i^{r+s}t^r u^s/(r!s!).$$

We refer to $\kappa_{rs}$ as the cumulant of order (r,s). The cumulants are given by

$$\kappa_{10} = EX = \mu_1,\ \kappa_{01} = EY = \mu_2,$$

$$\kappa_{20} = \mathrm{Var}(X) = \gamma_{11} + \int x^2(x^2 + y^2)^{-1}M(dx,dy),$$

$$\kappa_{02} = \mathrm{Var}(Y) = \gamma_{22} + \int y^2(x^2 + y^2)^{-1}M(dx,dy),$$

$$\kappa_{11} = \mathrm{Cov}(X,Y) = \gamma_{12} + \int xy(x^2 + y^2)^{-1}M(dx,dy),$$

and, for $r + s > 3$,

$$(2.2) \qquad \kappa_{rs} = \int x^r y^s(x^2 + y^2)^{-1}M(dx,dy).$$

(These formulas are readily obtained: When the cumulant of order (r,s) exists, it is the mixed partial of order (r,s), evaluated at (t,u) = (0,0); in this case, the differentiation may be passed under the integral sign.)

If $\underline{X}$ has no Gaussian component, then the cumulants are given by

$$\kappa_{10} = \mu_1, \ \kappa_{01} = \mu_2,$$

and, for $r + s \geq 2$,

$$\kappa_{rs} = \int x^r y^s (x^2+y^2)^{-1} M(dx,dy).$$

In the theory of multivariate inf.div. distributions the cumulant of order (2,2), $\kappa_{22}$, plays a special role; following [9], denote this functional by $\pi(X,Y)$.

As is well known [1,p.39], if $(X,Y)'$ is bivariate normal, then $\kappa_{rs} = 0$ if $r + s > 2$; in particular, then, $\pi(X,Y)=0$ if $(X,Y)'$ is bivariate normal.

Now suppose that a r.vec. $(U,V)'$ has the property that $(U,V) \equiv (U_1,V_1) + \ldots + (U_m,V_m)$, where the pairs $(U_1,V_1)$, $(U_2,V_2)$,..., $(U_m,V_m)$ are independent. Then (introducing obvious notation)

$$\sum \sum \kappa_{rs}(U,V) i^{r+s} t^r u^s/(r!s!) = \psi_{U,V}(t,u) = \log E\exp[i(tU+uV)]$$

$$= \log E\exp[i(t\textstyle\sum U_j + u\sum V_j)] = \log E \ \Pi \exp[i(tU_j + uV_j)]$$

$$= \log \Pi \ E \exp[i(tU_j + uV_j)] = \log \Pi \ \phi_{U_j,V_j}(t,u)$$

$$= \textstyle\sum \log \phi_{U_j,V_j}(t,u) = \sum \psi_{U_j,V_j}(t,u)$$

$$= \textstyle\sum_j \sum_r \sum_s \kappa_{rs}(U_j,V_j) i^{r+s} t^r u^s/(r!s!)$$

$$= \textstyle\sum_r \sum_s [\sum_j \kappa_{rs}(U_j,V_j)] i^{r+s} t^r u^s/(r!s!)$$

so that

$$\kappa_{rs}(U,V) = \textstyle\sum_j \kappa_{rs}(U_j,V_j).$$

Now, $(X,Y) \equiv (X_G,Y_G) + (X_P,Y_P)$, where the summands are independent, so, if the fourth moments are finite, so that $\pi(X,Y)$ exists,

$$\begin{aligned}\pi(X,Y) &= \kappa_{22}(X,Y) \\ &= \kappa_{22}(X_G,Y_G) + \kappa_{22}(X_P,Y_P) \\ &= \quad 0 \quad + \pi(X_P,Y_P) = \pi(X_P,Y_P).\end{aligned}$$

Note also that, if $(X,Y)'$ is inf.div. and the fourth moments are finite, then

$$(2.3) \qquad \pi(X,Y) = \kappa_{22}(X,Y) = \psi_{ttuu}(0,0) = \int x^2y^2(x^2 + y^2)^{-1}M(dx,dy) \geq 0.$$

This proves

THEOREM 2.1. If $(X,Y)'$ is inf.div. with finite fourth moments, then $\pi(X,Y) \geq 0$.

## 2.3. Independence

Now let us examine the relationship between independence of jointly inf.div. r.v.'s X,Y and nullity of the functionals Cov(X,Y) and $\pi(X,Y)$.

Note from (2.3) that $\pi(X,Y) = 0$ implies that M puts mass only on the axes of the (x,y)-plane, i.e., $X_P$ and $Y_P$ are independent. This proves

THEOREM 2.2. Let $(X,Y)'$ be inf.div. with finite fourth moments and no Gaussian component. Then X and Y are independent if and only if (iff.) $\pi(X,Y) = 0$.

In particular, discrete r.vec.'s can have no Gaussian component. Hence we have

COROLLARY 2.1. If $(X,Y)'$ is inf.div. with finite fourth moments and discrete, then X and Y are independent iff. $\pi(X,Y) = 0$.

THEOREM 2.3. Let $(X,Y)'$ be an inf.div. r.vec. Then X and Y are independent if $\mathrm{Cov}(X,Y) = 0$ and $\pi(X,Y) = 0$.

PROOF. We have $(X,Y) \equiv (X_G,Y_G) + (X_P,Y_P)$, where $\underline{X}_G$ and $\underline{X}_P$ are independent. To show independence of X and Y, it suffices to show $X_P,Y_P$ independent and $X_G,Y_G$ independent. Now,

(i) $0 = \pi(X,Y) = \kappa_{22}(X,Y) = \kappa_{22}(X_G,Y_G) + \kappa_{22}(X_P,Y_P) = 0 + \kappa_{22}(X_P,Y_P)$,

i.e., $\pi(X_P,Y_P) = 0$, which by Theorem 2.2 implies that $X_P$ and $Y_P$ are independent. Then,

$$
\begin{aligned}
\text{(ii) } 0 &= \mathrm{Cov}(X,Y) \\
&= \mathrm{Cov}(X_G+X_P, Y_G+Y_P) \\
&= \mathrm{Cov}(X_G,Y_G) + \mathrm{Cov}(X_P,Y_P) + \mathrm{Cov}(X_G,Y_P) + \mathrm{Cov}(X_P,Y_G) \\
&= \mathrm{Cov}(X_G,Y_G) + \mathrm{Cov}(X_P,Y_P) + 0 + 0 \quad \text{(by independence of } \underline{X}_G \text{ and } \underline{X}_P\text{)} \\
&= \mathrm{Cov}(X_G,Y_G) + 0 \qquad \text{(by (i))} \\
&= \mathrm{Cov}(X_G,Y_G).
\end{aligned}
$$

Therefore, $\mathrm{Cov}(X_G,Y_G) = 0$, and $X_G$ and $Y_G$ are independent because they are jointly normally distributed.

If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$ and also the cross-cumulants are null; in particular, $\pi(X,Y) = 0$. Together with Theorem 2.3 this gives

COROLLARY 2.2. Let $(X,Y)'$ be an inf.div. r.vec. Then X and Y are independent iff. $\mathrm{Cov}(X,Y) = 0$ and $\pi(X,Y) = 0$.

Now,

$$\text{(2.4)} \qquad \mathrm{Cov}\left[(X-EX)^2, (Y-EY)^2\right] = 2\left[\mathrm{Cov}(X,Y)\right]^2 + \kappa_{22}(X,Y).$$

Write (2.4) as $\tau(X,Y) = \nu(X,Y) + \pi(X,Y)$, where $\tau(X,Y) = \mathrm{Cov}\left[(X-EX)^2, (Y-EY)^2\right]$

and $\nu(X,Y) = 2[\mathrm{Cov}(X,Y)]^2$.

Here one may think of $\tau$ (or t) as "total," $\nu$ (or n) as "normal," and $\pi$ (or p) as "Poisson-type." Thus we have

LEMMA 2.1. Let (X,Y) be an inf.div. r.vec. Then $\mathrm{Cov}(X,Y) = 0$ and $\pi(X,Y) = 0$ implies $\mathrm{Cov}[(X-EX)^2, (Y-EY)^2] = 0$.

LEMMA 2.2. Let (X,Y)' be an inf.div. r.vec. Then $\mathrm{Cov}[(X - EX)^2, (Y-EY)^2 = 0$ implies $\mathrm{Cov}(X,Y) = 0$ and $\pi(X,Y) = 0$.

PROOF. For inf.div. r.vec.'s, $\pi(X,Y) \geq 0$. Hence $0 = \mathrm{Cov}[(X-EX)^2, (Y-EY)^2] = 2[\mathrm{Cov}(X,Y)]^2 + \pi(X,Y)$ implies $\mathrm{Cov}(X,Y) = 0$ and $\pi(X,Y) = 0$.

COROLLARY 2.3. Let (X,Y)' be an inf.div. r.vec. Then nullity of both Cov(X,Y) and $\pi(X,Y)$ is equivalent to nullity of $\mathrm{Cov}[(X-EX)^2, (Y-EY)^2]$.

THEOREM 2.4. Let (X,Y)' be an inf.div. r.vec. Then X and Y are independent iff. $\mathrm{Cov}[X-EX)^2, (Y-EY)^2] = 0$.

PROOF. If (X,Y)' is inf.div., then X and Y are independent iff. $\mathrm{Cov}(X,Y) = 0$ and $\pi(X,Y) = 0$, and this holds iff. $\mathrm{Cov}[(X-EX)^2, (Y-EY)^2] = 0$.

If we know that (X,Y)' is bivariate normal, then X and Y are independent iff. $\mathrm{Cov}(X,Y) = 0$. If we know only that (X,Y)' is inf.div. (and not necessarily bivariate normal), then we can be assured of independence of X and Y only if $\pi(X,Y) = 0$ as well as $\mathrm{Cov}(X,Y) = 0$.

## 2.4. Measures of dependence

Note that $\tau(X,Y)$ provides a measure of dependence between X and Y, in the sense that the correlation

$$\mathrm{Corr}[(X-EX)^2, (Y-EY)^2] =$$

$$\mathrm{Cov}[(X-EX)^2, (Y-EY)^2]/\{\mathrm{Var}[(X-EX)^2][(Y-EY)^2]\}^{1/2}$$

is at most one in absolute value. The meaning of a value of one is that there exist constants a and b such that $(Y-EY)^2 = a + b(X-EX)^2$ with probability one.

Similarly, $\pi(X,Y)$ provides a measure of dependence between X and Y when $(X,Y)'$ is inf.div., in the sense that [9]

$$(2.5) \qquad \pi(X,Y)/[\pi(X,X)\pi(Y,Y)]^{1/2}$$

is at most one (and at least zero, since $\pi(X,Y) > 0$ if $(X,Y)'$ is inf.div.) The meaning of a value of one for (2.5) when $(X,Y)'$ has no normal component is [9] that there exist independent r.v.'s U and V and a constant r such that $X-EX \equiv U+V$, $Y-EY \equiv r(U-V)$, where $r = [\pi(Y,Y)/\pi(X,X)]^{1/4}$. If $(X,Y)'$ has a normal component, then, when (2.5) has a value of one, $(X-EX,Y-EY)$ is distributed as $(Z_1+U+V, Z_2 + r(U-V))$, where $(Z_1,Z_2)$ is bivariate normal.

## 3. SEVERAL VARIABLES

THEOREM 3.1. Let $\underline{X} = (X_1,X_2,\ldots,X_p)'$ be an inf.div. r.vec.

(a) The variates of the Poisson component of $\underline{X}$ are independent iff.

$$(3.1) \qquad \pi(X_u,X_v) = 0 \text{ for all } u,v \text{ such that } u \neq v.$$

(b) The variates of $\underline{X}$ are independent iff. $\pi(X_u,X_v) = 0$ and $\nu(X_u,X_v) = 0$ for all u,v such that $u \neq v$.

(c) The variates of $\underline{X}$ are independent iff. $\tau(X_u,X_v) = 0$ for all u,v such that $u \neq v$.

REMARKS. (i) Note that, in particular, pairwise independence implies mutual independence in the multivariate inf.div. class. (See also Theorem 3.2 below.) (ii) The proof given below is essentially the same as that in [9] but is given here because it is important for the understanding of Remark (i), which in turn is important for the development of inference for multivariate inf.div. distributions.

PROOF. (a) The Kolmogorov representation for the log c.f. in the case of finite variances is (1.1). If $X_v$, $v = 1, 2,\ldots, p$, are independent, then, in particular, $X_u$ and $X_v$ are independent, and cross-cumulants are zero; in particular, $\kappa_{22}(X_u,X_v) = \pi(X_u,X_v) = 0$. If,

conversely, $\pi(X_u,X_v) = 0$ for all u,v such that $u \neq v$, then we have

$$0 = \sum_{\substack{u,v \\ u\neq v}}\sum \pi(X_u,X_v) = \int \sum_{\substack{u,v \\ u\neq v}}\sum x_u^2 \; x_v^2(\underline{x}'\underline{x})^{-1}M(d\underline{x}).$$

This implies that

$$\sum_{\substack{u,v \\ u\neq v}}\sum x_u^2 \; x_v^2 = 0 \tag{3.2}$$

at points $(x_1,x_2,\ldots,x_p)$ where M has mass. But (3.2) is true only for points $(x_1,x_2,\ldots,x_p)$ with at most one non-zero co-ordinate, i.e., only on the axes of $R^p$. Therefore, M can have mass only on the axes. Hence, $M = M_1 + \ldots + M_p$, where $M_v$ has mass only on the $x_v$-axis. Let $h(u_1,u_2,\ldots,u_p;\underline{x})$ be the integrand of (1.1). Then the log c.f. of the Poisson component is

$$\psi(\underline{u}) + \underline{u}'\underline{\Gamma}\;\underline{u} - i\underline{u}'\underline{\mu} = \int h(u_1,u_2,..,u_p;\underline{x}) \sum_{v=1}^{p} M_v(dx)$$

$$= \sum_{v=1}^{p} \int h(0,\ldots,0,u_v,0,\ldots,0;\underline{x})M_v(d\underline{x})$$

$$= \sum_{v=1}^{p} \int h(0,\ldots,0,u_v,0,\ldots,0)M(d\underline{x})$$

$$= \sum_{v=1}^{p} \psi_v(u_v),$$

where $\psi_v(u_v)$ is the log c.f. of an inf.div. variate. Thus the Poisson component of $\underline{X}$ has independent variates.

(b) We have $\underline{X} \equiv \underline{X}_G + \underline{X}_p$. where $\underline{X}_G$ and $\underline{X}_p$ are independent. It suffices to show that the variates of $\underline{X}_p$ are independent and the variates of $\underline{X}_G$ are independent. Now, (3.1) gives independence of the variates of $\underline{X}_p$.

Further, letting $\underline{X}_G = (X_{G1}, X_{G2}, \ldots, X_{Gp})$ and similarly for $\underline{X}_p$, we have $0 = \mathrm{Cov}(X_u, X_v) = \mathrm{Cov}(X_{Gu} + X_{Pu}, X_{Gv} + X_{Pv})$, which, as in the proof of Theorem 2.3, gives independence of $X_{Gu}$ and $X_{Gv}$. Thus the variates of $\underline{X}_G$ are uncorrelated and jointly normal, and hence independent.

(c) We have $\tau(X_u, X_v) = \mathrm{Cov}[(X_u - EX_u)^2, (X_v - EX_v)^2] = \nu(X_u, X_v) + \pi(X_u, X_v)$. Further, $\pi(X_u, X_v) \geq 0$. Hence, $\tau = 0$ iff. $\nu = 0$ and $\pi = 0$, i.e., iff. $\pi = 0$ and $\mathrm{Cov}(X_u, X_v) = 0$.

<u>THEOREM 3.2</u>. Let $\underline{X}$ be an inf.div. r.vec. Then pairwise independence of the elements of $\underline{X}$ is equivalent to their mutual independence.

<u>REMARKS</u>. (i) The result in one direction is trivial: mutual independence is stronger than pairwise independence. (ii) Now suppose the variates are pairwise independent. Theorem 3.1 gives the result in the case of finite fourth moments. The result is proved in general, without the assumption of fourth moments, in [5]. In doing this, the Levy representation (see, e.g., [12, p. 160], for a discussion of the univariate case) was used. It is of the form

$$\psi(\underline{u}) = i\underline{u}'\underline{\alpha} - \underline{u}'\underline{\Gamma}\,\underline{u}/2 + \int[\exp(i\underline{u}'\underline{x}) - 1 - i\underline{u}'\underline{x}/(1 + \underline{x}'\underline{x})]L(dx)$$

and exists even when the variances are not finite.

## 4. SOME REMARKS ON INFERENCE

Consider the decomposition $\underline{X} \equiv \underline{X}_G + \underline{X}_p$ and the analogous decomposition $\underline{T} = \underline{N} + \underline{\Pi}$, where $\underline{T} = [\tau(X_u, X_v)]$, $\underline{N} = [\nu(X_u, X_v)]$, $\underline{\Pi} = [\pi(X_u, X_v)]$. In general, one must consider estimates of both $\underline{N}$ and $\underline{\Pi}$. However, if $\underline{X}$ is discrete it has no Gaussian component and one can consider $\underline{\Pi}$ alone. Use of estimates of the parameters $\pi(X_u, X_v)$ in analyzing a set of data is illustrated in [11]. A formula for an unbiased estimate of $\pi(X_u, X_v)$, which is the cumulant of order (2,2), is given in a list of bivariate k-statistics in [5, p. 329, expression (13.2)].

### <u>4.1. Testing Independence</u>

In [10] a test of independence of X and Y, where (X,Y)' is an

inf.div. r.vec., is given. The test is based on the asymptotic normality of a sample analog of $\tau(X,Y)$. That is, it is based on the ratio of a sample analog of this parameter to its asymptotic standard error. This statistic is treated as a normal deviate. Similarly, in [10] a test of independence of $X_1,X_2,\ldots,X_p$, assuming $\underline{X}$ is inf.div. is based on the asymptotic chi-square distribution of a suitable quadratic form in the statistics $t_{uv}$, $u,v = 1,2,\ldots p$, $u < v$, where $t_{uv}$ is a sample analog of $\tau(X_u,X_v)$.

## 4.2. Testing Normality

In the continuous case one would want to test whether he could restrict attention to the normal family, versus using the full inf.div. class. It was noticed in [2] and later independently in [9] that nullity of the fourth cumulant characterizes the normal distribution in the class of inf.div. distributions. Similarly [9], nullity of the fourth cumulants of a multivariate distribution characterizes it as multivariate normal in the class of multivariate inf.div. distributions.

Multivariate goodness-of-fit problems seem to be rather difficult, and the problem of testing multivariate normality has received a fair amount of attention; see [3], [7], [8]. Sometimes such testing problems are considerably simplified when the class of alternatives is reduced. When the class is reduced to the inf.div. laws, still a large class, the resulting testing problem is shown to be quite tractable, due to the simple characterization of the normal family in the inf.div. class.

As remarked above, [10] shows how to test independence of the variates of an inf.div. r.vec. The test given here can be used as a preliminary test to decide whether one wants to use the full inf.div. model or rather to rely on the normal model. In the normal model tests of independence are of course based on the correlation coefficients, or, equivalently, on the covariances. In the full inf.div. model, tests of independence are based on the covariances of the squares of the centered

variables [10].

Given any random vector $\underline{X}$, let $L(\underline{X})$ denote the probability law of $\underline{X}$. Let C denote the class of inf.div. laws with finite eighth moments (see below). Let N denote the family of normal laws. We wish to test the hypothesis H: $L(\underline{X})$ is in N against alternatives that $L(\underline{X})$ is in C - N.

The hypothesis test. Recall that for a normal distribution $\kappa_r = 0$ for $r > 3$. The hypothesis test is based on the following characterization of multivariate normality in the inf.div. class, due to [9].

THEOREM 4.1. Let $\underline{X}$ be an inf.div. r.vec. with variates $X_v$, $v = 1,2,\ldots,p$, which have finite fourth moments. Then $\underline{X}$ is distributed according to a multivariate normal distribution iff. $\kappa_4(X_v) = 0$, for $v = 1,2,\ldots,p$.

REMARK. Given a r.v. Y, let $\mu_r(Y)$ denote its r-th central moment. Then $\kappa_4(Y) = \mu_4(Y) - 3[\mu_2(Y)]^2$, so that $\kappa_4/\mu_2$ is a conventional measure of kurtosis. [6, p. 88].

An unbiased estimate of $\kappa_4$ is the fourth k-statistic [6, formulas (12.28) and (12.29), pp. 299-300], viz.,

$$k_4 = [(n^3+ n^2)s_4 - 4(n^2+ n)s_3s_1 - 3(n^2-n)s_2^2 + 12n\ s_2s_1^2 - 6s_1^4]/n^{[4]}$$

$$= n^2[(n+1)m_4 - 3(n-1)m_2^2]/(n-1)^{[3]},$$

where, for a univariate sample $y_1,y_2,\ldots,y_n$,

$$s_r = \sum_{i=1}^{n} y_i^r, \quad m_r = \sum_{i=1}^{n} (y_i - \bar{y})^r/n, \quad \bar{y} = \sum_{i=1}^{n} y_i/n,$$

and

$$n^{[r]} = n(n-1)\ldots(n-r+1).$$

Let

$$\underline{\kappa} = (\kappa_4(X_1), \ldots, \kappa_4(X_p))'.$$

The hypothesis H is equivalent to $\underline{\kappa} = \underline{0}$. Let $\underline{k} = (k_4(X_1), \ldots, k_4(X_p))'$. Let $\underline{\Sigma}(\underline{k})$ denote the covariance matrix of $\underline{k}$. Then, by the asymptotic

joint normality of functions of sample moments, the quadratic form $(\underline{k}-\underline{\kappa})'[\underline{\Sigma}(\underline{k})]^{-1}(\underline{k}-\underline{\kappa})$ is asymptotically distributed according to the chi-square distribution with p degrees of freedom (d.f.). Under the hypothesis the quadratic form $\underline{k}'([\underline{\Sigma}(\underline{k})]^{-1}\underline{k}$ is asymptotically distributed according to chi-square with p d.f. Further, if $\underline{S}(\underline{k})$ is a consistent estimate of $\underline{\Sigma}(\underline{k})$, then quadratic form $Q = \underline{k}'[\underline{S}(\underline{k})]^{-1}\underline{k}$ is under H also asymptotically distributed as chi-square with p d.f. To construct a test it suffices that $\underline{S}(\underline{k})$ be consistent for $\underline{\Sigma}(\underline{k})$ under the hypothesis.

The covariance of $k_4(X)$ and $k_4(Y)$ is obtainable from [6, formula (13.58), p. 343], as

$$\begin{aligned}\mathrm{Cov}[k_4(X),k_4(Y)] &= \kappa_{44}/n + 16\kappa_{11}\kappa_{33}/n(n-1) + 48\kappa_{21}\kappa_{23}/(n-1)\\ &+ 72n\kappa_{11}^2\kappa_{22}/(n-1)^{[2]} + (16\kappa_{13}\kappa_{31} + 18\kappa_{22}^2)/(n-1)\\ &+ 144n\kappa_{11}\kappa_{21}\kappa_{12}/(n-1)^{[2]} + 24n(n+1)\kappa_{11}^4/(n-1)^{[3]}.\end{aligned}$$

For normal distributions $\kappa_{rs} = 0$ if $r+s > 3$, so the covariance above reduces to

$$24n(n+1)\kappa_{11}^4/[(n-1)(n-2)(n-3)];$$

also,

$$\mathrm{Var}(k_4) = 24n(n+1)\kappa_2^4/[(n-1)(n-2)(n-3)].$$

Thus we can take the (u,v)-th element of $\underline{S}(\underline{k})$ to be

$$24n(n+1)\ s_{uv}^4/[(n-1)(n-2)(n-3)],$$

where $s_{uv}$ is the usual sample covariance,

$$s_{uv} = \sum_{i=1}^{n} (X_{ui}-\bar{X}_u)(X_{vi}-\bar{X}_v)/(n-1), \quad u, v = 1,2,\ldots,p,$$

where $\underline{X}_i = (X_{1i},X_{2i},\ldots,X_{pi})'$, $i = 1,2,\ldots,n$, and $\bar{X}_v = \sum_{i=1}^{n} X_{vi}/n$.

At level $\alpha$ one rejects the hypothesis of normality if Q exceeds the upper $\alpha$-th percentage point of the chi-square distribution with p d.f.

Example. The procedure is illustrated for the Fisher iris data. Fifty observations were taken from each of the populations, *Iris setosa*, *versicolor*, and *virginica*. The variables are $x$ = sepal length, $x$ = sepal width, $x$ = petal length, and $x$ = petal width. The table gives the values of the coefficient of kurtosis, $b_2 = k_4/m_2$, where $k_4$ and $m_2$ are as defined above. The variance of $b_2$ for a normal parent is approximately $24/n$. Since here $n = 50$, the standard error of any one $b_2$ in the table is approximately 0.693. One could use this to test the significance of any one of the twelve $b_2$'s given in the table, but then one would run into the multiple-comparisons problem. The multivariate test avoids this problem. The chi-square values in the table are values of quadratic forms in the individual $k_4$'s for the four separate variables and hence may be considered as quadratic forms in the $b_2$'s for the four separate variables. For *Iris setosa* the chi-square value is almost significant at the 5% level. For the other two species, the values are not significant. Adding the three independent chi-squares, we have $9.390 + 1.167 + 1.770 = 12.327$, with 12 d.f.; $P = .43$.

COEFFICIENTS OF KURTOSIS
FOR THE FOUR VARIABLES OF THE FISHER IRIS DATA,
WITH VALUES OF CHI-SQUARE FOR TESTING NORMALITY
AGAINST INFINITELY DIVISIBLE ALTERNATIVES

| | *Iris setosa* | *Iris versicolor* | *Iris virginica* |
|---|---|---|---|
| Values of coefficients of kurtosis: | | | |
| Variable 1 | -0.263 | -0.555 | 0.034 |
| Variable 2 | 0.994 | -0.381 | 0.735 |
| Variable 3 | 1.064 | 0.050 | -0.160 |
| Variable 4 | 1.790 | -0.427 | -0.627 |
| Values of chi-square test statistic (4 d.f.): | | | |
| | 9.390 (P=.052) | 1.167 (P=.88) | 1.770 (P=.78) |

ACKNOWLEDGEMENTS

Parts of the research reported here were supported under Grants AFOSR 76-3050 and AFOSR 77-3454 from the Air Force Office of Scientific Research and Contract N00014-80-C-0408, Task NR042-443 from the Office of Naval Research.

## REFERENCES

[1] Anderson, T.W.: An Introduction to Multivariate Statistical Analysis. New York, Wiley (1958).

[2] Borges, R.: A characterization of the normal distribution (a note a paper by Kozin). Z. Wahrsch. verw. Gebiete 5, 244-246 (1966).

[3] Cox, D.R., and Small, N.J.H.: Testing multivariate normality. Biometrika 65, 263-272 (1978).

[4] Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugenics 7, 179-188 (1936). Reprinted in Fisher, R.A.: Contributions to Mathematical Statistics, New York, Wiley (1950).

[5] Hudson, W.N., and Tucker, H.G.: Asymptotic independence in the multivariate central limit theorem. Ann. Probab. 7, 662-671 (1979).

[6] Kendall, M.G., and Stuart, A.: The Advanced Theory of Statistics. Vol. 1: Distribution Theory (4th ed.). New York, Hafner (Macmillan); London, Griffin (1977).

[7] Malkovich, J.F., and Afifi, A.A.: On tests for multivariate normality. J. Amer. Statist. Assn. 68, 176-179 (1973).

[8] Mardia, K.V.: Applications of some measures of multivariate skewness and kurtosis to testing normality and to robustness studies. Sankhya(B) 36, 115-128 (1974).

[9] Pierre, P.: Infinitely divisible distributions, conditions for independence, and central limit theorems. J. Math. Anal. Appl. 33, 341-354 (1971).

[10] Sclove, S.L.: Testing independence of variates in an infinitely divisible random vector. J. Multivariate Anal. 8, 479-485 (1978).

[11] Sclove, S.L.: Modeling the distribution of fingerprint characteristics. In C. Taillie, G.P. Patil, B. Baldessari (eds.), Statistical Distributions in Scientific Work 6, 111-130, Dordrecht, Holland, D. Reidel (1981).

[12] Tucker, H.G.: A Graduate Course in Probability, New York, Academic Press (1967).

Department of Quantitative Methods
College of Business Administration
University of Illinois at Chicago Circle
P.O. Box 4348
Chicago, Illinois 60680
U.S.A.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ONR Technical Report 82-3

2. GOVT ACCESSION NO.: AD-A116382

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Some Aspects of Inference for Multivariate Infinitely Divisible Distributions

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Stanley L. Sclove

8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0408

9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Illinois at Chicago Circle, Box 4348, Chicago, IL 60680

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:

12. REPORT DATE: June 15, 1982

13. NUMBER OF PAGES: 17 + ii

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Statistics & Probability Branch, Office of Naval Research, Department of the Navy, Arlington, VA 22217

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Unlimited distribution

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

infinite divisibility; multivariate normality; kurtosis; measures of dependence

American Mathematical Society (AMS) classification: 62H15; 62H20

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Measurement of dependence in the infinitely divisible class of multivariate distributions, based on developments in probability theory for that class, is discussed. It has been shown that pairwise independence is equivalent to mutual independence in the infinitely divisible class. When the infinitely divisible variables contain no normal component (in particular, when they are discrete), the cumulant of order (2,2) can be used as a measure of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

pairwise dependence; when a normal component is present, the appropriate measure of pairwise dependence also involves the covariance. Results for testing independence of infinitely divisible random variables are discussed. A method of testing normality against infinitely divisible alternatives is given.